

# Distribution of Quantiles in Samples from a Bivariate Population\*

M. M. Siddiqui

(March 3, 1960)

Let  $F(x, y)$  be the joint distribution function of  $(X, Y)$ , possessing a probability density function  $f(x, y)$ . Let  $F_1(x)$  and  $F_2(y)$  be the marginal distribution functions of  $X$  and  $Y$  respectively. Let  $\alpha$  be a quantile of  $F_1(x)$  and  $\beta$  be a quantile of  $F_2(y)$ . A random sample  $(X_k, Y_k)$ ,  $k=1, 2, \dots, n$ , is drawn and the values on each variate are ordered so that  $X'_i \leq X'_j$  and  $Y'_i \leq Y'_j$  if  $i \leq j$ . Let  $i$  and  $j$  be the greatest integers such that  $i/n \leq F_1(\alpha)$  and  $j/n \leq F_2(\beta)$ , and let  $M$  be the number of elements  $(X, Y)$  such that  $X < X'_i$  and  $Y < Y'_j$ . The joint distribution of  $(M, X'_i, Y'_j)$  is obtained and is shown to be asymptotically normal. Estimates and confidence limits on the parameters of interest are also given.

## 1. Summary

The exact distribution of quantiles as well as its limiting form is well known in the univariate case.<sup>1</sup> Mood<sup>2</sup> investigated the joint distribution of medians in samples from a multivariate population and showed that it is asymptotically normal. In this paper the exact joint distribution of quantiles and an auxiliary statistic in samples from a bivariate population are obtained and it is shown that the joint distribution is asymptotically trivariate normal. The auxiliary statistic is utilized to estimate the correlation coefficient between quantiles and also for setting up confidence limits on it. Estimates for an ordinate of a univariate probability density function (pdf) are also obtained. These are used in setting confidence limits on the quantiles.

## 2. Introduction

Let  $F(x, y)$  be the distribution function of  $(X, Y)$  possessing a pdf  $f(x, y)$ . Let the marginal distribution function of  $X$  be denoted by  $F_1(x)$  and that of  $Y$  by  $F_2(y)$ . Let  $\alpha$  be a quantile of  $F_1(x)$  and  $\beta$  be a quantile of  $F_2(y)$ . It is assumed that the first and second partial derivatives of  $F(x, y)$  are continuous in a neighborhood of  $(\alpha, \beta)$  and that  $f(\alpha, \beta) \neq 0$ . A random sample  $(X_k, Y_k)$ ,  $k=1, 2, \dots, n$ , is drawn from  $F(x, y)$ . Let the sample values of  $X$  be ordered so that

$$X'_1 < X'_2 < \dots < X'_n.$$

Similarly, let  $Y'_1 < Y'_2 < \dots < Y'_n$ . All such samples for which  $X'_i = X'_j$  or  $Y'_i = Y'_j$  for different indices  $i$  and  $j$  may be excluded from consideration, as these samples form a set of probability zero. Let  $i$  and  $j$  be the integers such that

$$i/n \leq F_1(\alpha) < (i+1)/n; \quad j/n \leq F_2(\beta) < (j+1)/n.$$

Hence  $i/n = F_1(\alpha) - \delta_1/n$ ,  $j/n = F_2(\beta) - \delta_2/n$ ,  $0 \leq \delta_1, \delta_2 < 1$ .

If  $M$  denotes the number of elements  $(X, Y)$  in the sample such that  $X < X'_i$  and  $Y < Y'_j$ , then  $M$  is a discrete-valued random variable with possible values  $0, 1, \dots, \min(i-1, j-1)$ .

First the exact distribution of  $(M, X'_i, Y'_j)$  is obtained for a fixed  $n$ . Then the asymptotic distribution of the standardized variates is found when  $n, i, j \rightarrow \infty$  so that  $i/n \rightarrow F_1(\alpha)$  and  $j/n \rightarrow F_2(\beta)$ .

## 3. Joint Distribution of the Three Variables

Let us take a Euclidean plane  $(x, y)$  to represent the sample points. Given two numbers  $x_0$  and  $y_0$ , the lines  $x=x_0$  and  $y=y_0$  divide the plane into four quadrants, namely

$$Q_1 = \{(x, y): x < x_0, y < y_0\},$$

$$Q_2 = \{(x, y): x < x_0, y > y_0\},$$

$$Q_3 = \{(x, y): x > x_0, y < y_0\},$$

$$Q_4 = \{(x, y): x > x_0, y > y_0\}.$$

Let  $p_i = \Pr\{(X, Y) \in Q_i\}$ ,  $i=1, 2, 3, 4$ . Then

$$p_1 = F(x_0, y_0), \quad p_2 = F_1(x_0) - F(x_0, y_0),$$

$$p_3 = F_2(y_0) - F(x_0, y_0), \quad p_4 = 1 - F_1(x_0)$$

$$- F_2(y_0) + F(x_0, y_0).$$

$X'_i$  will fall in  $(x_0, x_0 + dx_0)$  and  $Y'_j$  will fall in  $(y_0, y_0 + dy_0)$  in all samples  $(X_k, Y_k)$ ,  $k=1, \dots, n$ , such that the following statements hold simultaneously:

- (1)  $i-1$  points fall on the left of the line  $x=x_0$ ,  $n-i$  points on the right of it, and one point on it;
- (2)  $j-1$  points fall below the line  $y=y_0$ ,  $n-j$  points above it, and one point on it.

Let  $(x_0, Y')$  and  $(X', y_0)$  denote the points which fall on  $x=x_0$  and  $y=y_0$  respectively. There are five distinct possibilities:

\*Contribution from the National Bureau of Standards Boulder Laboratories, Boulder, Colorado.

<sup>1</sup> Harald Cramér, Mathematical methods of statistics, p. 368 (Princeton Univ. Press, Princeton, N.J., 1946).

<sup>2</sup> A. M. Mood, On the joint distribution of medians in samples from a multivariate population, Ann. Math. Stat. **12**, 268 (1941).

- Case (1):  $X' < x_0$  and  $Y' < y_0$ ,  
Case (2):  $X' < x_0$  and  $Y' > y_0$ ,  
Case (3):  $X' > x_0$  and  $Y' < y_0$ ,  
Case (4):  $X' > x_0$  and  $Y' > y_0$ ,  
Case (5):  $X' = x_0$ , in which case there is only one point  $(x_0, y_0)$ , common to both the lines.

We will consider  $Pr(M=m, x_0 \leq X'_i \leq x_0 + dx_0, y_0 \leq Y'_i \leq y_0 + dy_0) = P(m, x_0, y_0) dx_0 dy_0$ . If  $M=m$ , then, in case (1), there will be  $i-m-2, j-m-2$  and  $n-i-j+m+2$  points in  $Q_2, Q_3$ , and  $Q_4$  respectively. The probability of such a sample is given by

$$P_1(m, x_0, y_0) dx_0 dy_0 = \frac{n! p_1^m p_2^{i-m-2} p_3^{j-m-2} p_4^{n+m-i-j+2} \frac{\partial F}{\partial x} \frac{\partial F}{\partial y} dx_0 dy_0}{m!(i-m-2)!(j-m-2)!(n+m-i-j+2)!}, \quad (3.1)$$

where the partial derivatives are evaluated at  $(x_0, y_0)$ . The contribution to  $P(m, x_0, y_0)$  from other cases may be found similarly. The possible values of  $m$  in each case are determined by the rule that negative factorials are not allowed. For example, in case (1), the maximum possible value of  $m$  is  $\min(i-2, j-2)$  and the minimum possible value is  $\max(0, i+j-n-2)$ . It will be assumed that the suffixes 1 and 2 are so chosen that  $F_1(\alpha) \leq F_2(\beta)$  and hence  $i \leq j$ . Thus, with the convention that the terms involving negative factorials be replaced with zeros,

$$P(m, x_0, y_0) = \sum_{k=1}^5 P_k(m, x_0, y_0), \quad m=0, 1, \dots, i-1, \quad -\infty \leq x_0, \quad y_0 \leq \infty, \quad (3.2)$$

where  $P_1$  is as given in (3.1) and

$$\left. \begin{aligned} P_2 &= \frac{n! p_1^m p_2^{i-m-2} p_3^{j-m-1} p_4^{n+m-i-j+1} \left( f_1(x_0) - \frac{\partial F}{\partial x} \right) \frac{\partial F}{\partial y}}{m!(i-m-2)!(j-m-1)!(n+m-i-j+1)!}, \\ P_3 &= \frac{n! p_1^m p_2^{i-m-1} p_3^{j-m-2} p_4^{n+m-i-j+1} \frac{\partial F}{\partial x} \left( f_2(y_0) - \frac{\partial F}{\partial y} \right)}{m!(i-m-1)!(j-m-2)!(n+m-i-j+1)!}, \\ P_4 &= \frac{n! p_1^m p_2^{i-m-1} p_3^{j-m-1} p_4^{n+m-i-j} \left( f_1(x_0) - \frac{\partial F}{\partial x} \right) \left( f_2(y_0) - \frac{\partial F}{\partial y} \right)}{m!(i-m-1)!(j-m-1)!(n+m-i-j)!}, \\ P_5 &= \frac{n! p_1^m p_2^{i-m-1} p_3^{j-m-1} p_4^{n+m-i-j+1} f(x_0, y_0)}{m!(i-m-1)!(j-m-1)!(n+m-i-j+1)!}, \end{aligned} \right\} \quad (3.3)$$

where all the partial derivatives are evaluated at  $(x_0, y_0)$ .

Hence the joint pdf of  $(X'_i, Y'_j)$ ,  $p(x_0, y_0)$ , is given by

$$p(x_0, y_0) = \sum_{m=0}^{i-1} P(m, x_0, y_0), \quad -\infty \leq x_0, \quad y_0 \leq \infty, \quad (3.4)$$

and the probability distribution of  $M$  is given by

$$P(m) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P(m, x_0, y_0) dx_0 dy_0, \quad m=0, 1, \dots, i-1. \quad (3.5)$$

## 4. Asymptotic Distribution

In this section the following will be found:

(i) The limit of the pdf of  $(U, V)$ ,  $p(u, v)$  (the notation  $p$  is used generically to denote a pdf) where

$$U = \frac{n^{1/2}(X'_i - \alpha) f_1(\alpha)}{[F_1(\alpha) \{1 - F_1(\alpha)\}]^{1/2}}, \quad V = \frac{n^{1/2}(Y'_j - \beta) f_2(\beta)}{[F_2(\beta) \{1 - F_2(\beta)\}]^{1/2}};$$

(ii) The asymptotic probability distribution of  $Q = M/n$  as  $n \rightarrow \infty$ .

The following operations are performed.

(1) Write

$$F(\alpha, \beta) = F, \quad F_1(\alpha) = F_1, \quad F_2(\beta) = F_2, \quad f_1(\alpha) = f_1,$$

$$f_2(\beta) = f_2, \quad \frac{\partial F}{\partial x} = g_1, \quad \frac{\partial F}{\partial y} = g_2, \quad \frac{\partial^2 F}{\partial x^2} = g_3, \quad \frac{\partial^2 F}{\partial y^2} = g_4,$$

$$\frac{\partial^2 F}{\partial x \partial y} = f, \quad g_1/f_1 = c_1, \quad g_2/f_2 = c_2, \quad f'_1(\alpha) = f'_1, \quad f'_2(\beta) = f'_2,$$

where all the partial derivatives are evaluated at  $(\alpha, \beta)$ . It is observed that by our assumptions none of the quantities  $f_1, f_2, g_1, g_2$ , and  $f$  is equal to zero. It is further observed that  $F \leq F_1$  since  $F(\alpha, \beta) \leq F(\alpha, \infty)$ . Also,

$$g_1 = \int_{-\infty}^{\beta} f(\alpha, y) dy \leq \int_{-\infty}^{\infty} f(\alpha, y) dy = f_1,$$

and  $g_2 \leq f_2$ , so that  $0 \leq c_1, c_2 \leq 1$ .

(2) For  $S \geq 2$ , use Stirling's approximation,

$$S! = \sqrt{2\pi} S^{S+1/2} e^{-S} [1 + O(S^{-1})].$$

(3) Write

$$m/n = q, \quad 1/n = dq.$$

This operation is equivalent to replacing the discrete random variable  $M/n$  by a continuous random variable  $Q$  over the range  $(0, F_1 - 1/n)$ . However,  $|Q - M/n| < 1/n$ .

If by these operations  $P(m, x_0, y_0) dx_0 dy_0$  is transformed into  $p(q, x_0, y_0) dq dx_0 dy_0$ , then

$$p(q, x_0, y_0) = [G(q, x_0, y_0)]^n H(q, x_0, y_0) [1 + O(n^{-1})], \quad (4.1)$$

where

$$G(q, x_0, y_0) = \frac{p_1^q p_2^{F_1 - q} p_3^{F_2 - q} p_4^{1 - F_1 - F_2 + q}}{q^q (F_1 - q)^{F_1 - q} (F_2 - q)^{F_2 - q} (1 - F_1 - F_2 + q)^{1 - F_1 - F_2 + q}}, \quad (4.2)$$

$$H(q, x_0, y_0) = \frac{n^{3/2} p_4^2 (F_1 - q)^{3/2} (F_2 - q)^{3/2}}{(2\pi)^{3/2} (p_2 p_3)^2 q^{1/2} (1 - F_1 - F_2 + q)^{5/2}} \left[ \frac{\partial F}{\partial x} \frac{\partial F}{\partial y} \right. \\ + p_3 p_4^{-1} (1 - F_1 - F_2 + q) (F_2 - q)^{-1} \left( f_1(x_0) - \frac{\partial F}{\partial x} \right) \frac{\partial F}{\partial y} \\ + p_2 p_4^{-1} (1 - F_1 - F_2 + q) (F_1 - q)^{-1} \frac{\partial F}{\partial x} \left( f_2(y_0) - \frac{\partial F}{\partial y} \right) \\ + p_2 p_3 p_4^{-2} (1 - F_1 - F_2 + q)^2 (F_1 - q)^{-1} (F_2 - q)^{-1} \\ \left. \left( f_1(x_0) - \frac{\partial F}{\partial x} \right) \left( f_2(y_0) - \frac{\partial F}{\partial y} \right) \right], \quad (4.3)$$

where all the partial derivatives are evaluated at  $(x_0, y_0)$ . The terms in  $H$  arise severally from  $P_1, P_2, P_3$ , and  $P_4$  after the factor  $G^n$  is taken out and are given in that order. The contribution from  $P_5$  is of order  $n^{-1}$  as compared to the other terms and hence absorbed in  $O(n^{-1})$  in (4.1).

(4) Make the transformations

$$u = \frac{n^{1/2} (x_0 - \alpha) f_1}{[F_1(1 - F_1)]^{1/2}}, \quad v = \frac{n^{1/2} (y_0 - \beta) f_2}{[F_2(1 - F_2)]^{1/2}},$$

so that

$$dx_0 dy_0 = \frac{\sqrt{F_1 F_2 (1 - F_1) (1 - F_2)}}{n f_1 f_2} du dv$$

and  $p(q, x_0, y_0) dx_0 dy_0 dq$  goes into  $p(q, u, v) dq du dv$ , where  $p$  is used generically to denote a pdf and is not the same function from equation to equation.

(5) Expand each function of  $(u, v)$  in Maclaurin's series up to the terms of order  $n^{-1}$ , e.g.,

$$p_1 = F + (x_0 - \alpha) g_1 + (y_0 - \beta) g_2 + \frac{(x_0 - \alpha)^2}{2} g_3 + \frac{(y_0 - \beta)^2}{2} g_4 \\ + (x_0 - \alpha)(y_0 - \beta) f + O(n^{-3/2}),$$

where  $x_0$  and  $y_0$  are replaced in terms of  $u$  and  $v$ . Then

$$\log G = w(q) + (x_0 - \alpha)(q - F) k_1 f_1 \{ k_2 c_1 - F(F_2 - F)(1 - F_2) \} \\ + (y_0 - \beta)(q - F) k_1 f_2 \{ k_2 c_2 - F(F_1 - F)(1 - F_1) \} \\ - \frac{1}{2} (x_0 - \alpha)^2 k_1 f_1^2 \{ k_2 c_1^2 + (1 - 2c_1)F(F_2 - F)(1 - F_2) \\ + (q - F)A_1 \} - \frac{1}{2} (y_0 - \beta)^2 k_1 f_2^2 \{ k_2 c_2^2 + (1 - 2c_2)F(F_1 - F)(1 - F_1) \\ + (q - F)A_2 \} - (x_0 - \alpha)(y_0 - \beta) k_1 f_1 f_2 \{ k_2 c_1 c_2 - c_1 F(F_1 - F)(1 - F_1) \\ - c_2 F(F_2 - F)(1 - F_2) + F(F_1 - F)(F_2 - F) + (q - F)A_3 \} + O(n^{-3/2}).$$

Here,  $c_1$  and  $c_2$  are defined in the beginning of this section;  $A_1, A_2$ , and  $A_3$  are some constants depending only on the parameters of  $F(x, y)$  and need not be specified, as it will be seen presently that  $(q - F)$  is of order  $n^{-1/2}$ ;

$$1/k_1 = F(F_1 - F)(F_2 - F)(1 - F_1 - F_2 + F),$$

$$k_2 = F_1 F_2 (1 - F_1 - F_2 + 2F) - F^2;$$

and

$$w(q) = q \log \frac{F}{q} + (F_1 - q) \log \frac{F_1 - F}{F_1 - q} + (F_2 - q) \log \frac{F_2 - F}{F_2 - q} \\ + (1 - F_1 - F_2 + q) \log \frac{1 - F_1 - F_2 + F}{1 - F_1 - F_2 + q}.$$

It is easy to verify that  $w(q)$  has an absolute maximum at  $q = F$  in the range of values of  $q$  and  $w(F) = w'(F) = 0$ . Expanding  $w(q)$  in a Taylor's series about  $q = F$ ,

$$w(q) = -\frac{1}{2} k_1 k_2 (q - F)^2 + O[(q - F)^3].$$

Each function of  $q$  is expanded about  $q=F$  and the transformation

$$t = \sqrt{k_1 k_2 n} (q - F) \quad (4.3)$$

is made with a similar transformation  $T$  on the random variable  $Q$ . The pdf of  $(T, U, V)$ ,  $p(t, u, v)$  is then given by

$$p(t, u, v) = (2\pi)^{-3/2} (1 - \rho^2)^{-1/2} \exp \left[ -\frac{1}{2} (t - u h_1 - v h_2)^2 - \frac{u^2 + v^2 - 2\rho uv}{2(1 - \rho^2)} \right] [1 + O(n^{-1/2})] \quad (4.4)$$

where

$$\rho = \frac{F - F_1 F_2}{\sqrt{F_1 F_2 (1 - F_1) (1 - F_2)}},$$

$$h_1 = \frac{\sqrt{k_1} [k_2 c_1 - F(F_2 - F)(1 - F_2)]}{\sqrt{(1 - \rho^2) F_2 (1 - F_2)}},$$

$$h_2 = \frac{\sqrt{k_1} [k_2 c_2 - F(F_1 - F)(1 - F_1)]}{\sqrt{(1 - \rho^2) F_1 (1 - F_1)}}. \quad (4.5)$$

The limits of variation of  $u$  and  $v$  are  $(-\infty, \infty)$  and those of  $t$  are  $(-d_1 \sqrt{n}, +d_2 \sqrt{n})$  where  $d_1$  and  $d_2$  are positive constants independent of  $n$ . In fact,

$$d_1 = \sqrt{k_1 k_2} \max(F, F_1 + F_2 - 1), \quad d_2 = \sqrt{k_1 k_2} (F_1 - F).$$

When  $n \rightarrow \infty$  the joint distribution of  $(T, U, V)$  is a trivariate normal with mean vector  $(0, 0, 0)$  and covariance matrix  $V$  obtained from

$$V^{-1} = \begin{bmatrix} 1 & -h_1 & -h_2 \\ -h_1 & h_1^2 + (1 - \rho^2)^{-1} & h_1 h_2 - \rho(1 - \rho^2)^{-1} \\ -h_2 & h_1 h_2 - \rho(1 - \rho^2)^{-1} & h_2^2 + (1 - \rho^2)^{-1} \end{bmatrix}. \quad (4.6)$$

Integrating out  $t$ , the asymptotic pdf of  $(U, V)$  is obtained as

$$p(u, v) = (2\pi)^{-1} (1 - \rho^2)^{-1/2} \exp \left[ -\frac{1}{2(1 - \rho^2)} \{u^2 + v^2 - 2\rho uv\} \right], \quad -\infty \leq u, v \leq \infty, \quad (4.7)$$

so that  $\rho$  is the correlation coefficient between  $U$  and  $V$  and hence between  $X'_i$  and  $Y'_j$ . The following theorem may then be stated:

**THEOREM.** *The asymptotic joint distribution of the variates  $(X'_i, Y'_j)$  is normal with parameters*

$$EX'_i = \alpha, \quad EY'_j = \beta,$$

$$\text{Var } X'_i = \frac{F_1(1 - F_1)}{n f_1^2}, \quad \text{Var } Y'_j = \frac{F_2(1 - F_2)}{n f_2^2},$$

$$\text{Cov}(X'_i, Y'_j) = \frac{F - F_1 F_2}{n f_1 f_2}.$$

Integrating  $p(t, u, v)$  with respect to  $u$  and  $v$ , it is found that the distribution of  $T$  is asymptotically normal with mean zero and variance

$$\sigma_T^2 = 1 + h_1^2 + h_2^2 + 2h_1 h_2 \rho.$$

Hence, from (4.3),  $Q$  is asymptotically normally distributed with mean  $F$  and variance

$$\sigma_Q^2 = (k_1 k_2 n)^{-1} (1 + h_1^2 + h_2^2 + 2h_1 h_2 \rho). \quad (4.8)$$

The variate  $M/n$ , which is discrete-valued, is distinguished from  $Q$ , which possesses a pdf. However, if  $q = m/n$ ,

$$\Pr(M/n \leq q) \cong \Pr(Q \leq q). \quad (4.9)$$

## 5. Confidence Limits on Quantiles

Since  $X'_i$  is asymptotically normal with mean  $\alpha$  and variance

$$\sigma_i^2 = F_1(1 - F_1)/(n f_1^2),$$

confidence limits on  $\alpha$  with confidence coefficient  $(1 - \gamma)$  are given by  $X'_i \pm Z_\gamma \sigma_i$ . Here,  $0 < \gamma < 1$ , and

$$\int_{Z_\gamma}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \frac{1}{2} \gamma.$$

However,  $\sigma_i$  remains undetermined unless  $f_1$  is known. Exact nonparametric confidence limits on quantiles are readily available.<sup>3</sup> However, an alternate procedure is outlined here which provides an asymptotic estimate of  $f_1$  and hence of  $\sigma_i$ .

Consider the joint distribution of  $(X'_{i-k}, X'_i, X'_{i+l})$ , where as  $n \rightarrow \infty$ ,  $i \rightarrow \infty$  in such a way that  $i/n \rightarrow F_1$  while  $k/n$  and  $l/n$  both tend to zero.  $k$  and  $l$  may be taken to be of order  $n^a$ ,  $0 \leq a < 1$ . The probability that  $t \leq X'_i \leq t + dt$ ,  $t_1 \leq X'_{i-k} \leq t_1 + dt_1$ ,  $t_2 \leq X'_{i+l} \leq t_2 + dt_2$ , is given by  $p(t, t_1, t_2) dt dt_1 dt_2$ , where

$$p(t, t_1, t_2) = \frac{n! [F_1(t_1)]^{i-k-1} [F_1(t) - F_1(t_1)]^{k-1} [F_1(t_2) - F_1(t)]^{l-1} [1 - F_1(t_2)]^{n-i-l} f_1(t) f_1(t_1) f_1(t_2)}{(i-k-1)! (k-1)! (l-1)! (n-i-l)!} \quad (5.1)$$

<sup>3</sup> S. S. Wilks, Order statistics, Am. Math. Soc. Bull. **54**, 6 (1948); G. E. Noether, On confidence limits for quantiles, Ann. Math. Stat. **19**, 416 (1948).

Make the following transformations:

$$U = \frac{\sqrt{n}(X'_i - \alpha)f_1}{\sqrt{F_1(1-F_1)}}, \quad U_1 = \frac{n}{k}(X'_i - X'_{i-k}),$$

$$U_2 = \frac{n}{l}(X'_{i+l} - X'_i), \quad (5.2)$$

with corresponding transformations from  $(t, t_1, t_2)$  to  $(u, u_1, u_2)$ . Use Stirling's approximations for factorials involving  $i$  and  $n$ . Expand each function of  $u$  in Maclaurin's series and let  $n \rightarrow \infty$ . We finally obtain the asymptotic joint pdf of  $(U, U_1, U_2)$  given by

$$p(u, u_1, u_2) = \frac{e^{-u^2/2}}{\sqrt{2\pi}} \frac{(kf_1)^k}{(k-1)!} u_1^{k-1} e^{-kf_1 u_1} \frac{(lf_1)^l}{(l-1)!} u_2^{l-1} e^{-lf_1 u_2}$$

$$-\infty \leq u \leq \infty, \quad 0 \leq u_1, \quad u_2 \leq \infty. \quad (5.3)$$

The following statements are true asymptotically.

(1)  $(U, U_1, U_2)$  form an independent set of variates.

(2)  $2kf_1 U_1 = 2nf_1(X'_i - X'_{i-k})$  is a  $\chi^2$  variate with  $2k$  degrees of freedom.

(3)  $2lf_1 U_2 = 2nf_1(X'_{i+l} - X'_i)$  is a  $\chi^2$  variate with  $2l$  degrees of freedom.

(4) Hence,

$$2n(X'_{i+l} - X'_{i-k})f_1$$

is a  $\chi^2$  variate with  $2(k+l)$  degrees of freedom, distributed independently of  $X'_i$ .

(5) Thus,

$$W = \frac{U(k+l)}{f_1(kU_1 + lU_2)} = \frac{(X'_i - \alpha)(k+l)}{\sqrt{nf_1(1-F_1)}(X'_{i+l} - X'_{i-k})} \quad (5.4)$$

is distributed as the ratio of a  $N(0,1)$  variate and

$$\left. \begin{aligned} \rho &= 4F - 1, & k_1^{-1} &= F^2(1/2 - F)^2, & k_2 &= F(1/2 - F), \\ h_1 &= \frac{2c_1 - 1}{\sqrt{1 - \rho^2}}, & h_2 &= \frac{2c_2 - 1}{\sqrt{1 - \rho^2}}, \\ \sigma_Q^2 &= \frac{F(1 - 2F)}{2n} + \frac{1}{16n} [(2c_1 - 1)^2 + (2c_2 - 1)^2 + 2\rho(2c_1 - 1)(2c_2 - 1)]. \end{aligned} \right\} \quad (6.1)$$

Now  $0 \leq c_1, c_2 \leq 1$ ,  $0 \leq F \leq 1/2$ , and the expression in the square brackets has a maximum equal to  $2 + 2|\rho| = \max \{8F, 4(1 - 2F)\}$  for variations of  $c_1$  and  $c_2$ . Thus,

$$\sigma_Q^2 \leq \frac{F(1 - 2F)}{2n} + \frac{\max(2F, 1 - 2F)}{4n} \leq 1/(4n). \quad (6.2)$$

The last bound is the least upper bound of the expression in the middle and is attained when  $F = 0$  or  $1/2$ .

Hence  $Q \pm Z_\gamma/(2\sqrt{n})$  are asymptotic nonparametric

an independent  $\chi^2$  variate divided by its number of degrees of freedom  $2(k+l)$ .

$W$  is independent of the parameter  $f_1$ ; however, the distribution function of such a ratio has not been tabulated, hence, at present, it is of little practical use. However,  $1/f_1$  can be estimated by

$$U_3 = (kU_1 + lU_2)/(k+l), \quad (5.5)$$

which has mean value equal to  $1/f_1$  and variance  $(k+l)^{-1}f_1^{-2}$ . Thus

$$s_i = \sqrt{F_1(1-F_1)}U_3/\sqrt{n} \quad (5.6)$$

is an unbiased estimator of  $\sigma_i$  with known distribution, still in the asymptotic sense.

In practice,  $k$  and  $l$  may be taken of order  $n^{1/2}$ .

## 6. Confidence Limits on $F$ and $\rho$

The asymptotic maximum likelihood estimator of  $F$  is  $M/n$  which has asymptotic variance  $\sigma_Q^2$  given by (4.8). Solution of the inequalities

$$-Z_\gamma \sigma_Q \leq Q - F \leq Z_\gamma \sigma_Q$$

gives

$$h_1(Q) \leq F \leq h_2(Q),$$

to obtain asymptotic confidence limits on  $F$  with confidence coefficient  $(1 - \gamma)$ . For large  $n$ ,  $\sigma_Q$  may be replaced by  $S_Q$ , where  $S_Q^2$  is obtained from the right-hand side of (4.8) by replacing  $F$  by  $Q$ . Thus, to order  $n^{-1/2}$ ,  $100(1 - \gamma)$  percent confidence limits for  $F$  are  $Q \pm Z_\gamma S_Q$ .

Here, again, there is the difficulty of nuisance parameters  $c_1$  and  $c_2$ . To show how to overcome this difficulty, the case of medians is presented.

When  $F_1 = F_2 = 1/2$ ,

confidence limits on  $F$  with confidence coefficient at least equal to  $1 - \gamma$ .

Going back to the general case, it is observed that

$$r = \frac{Q - F_1 F_2}{\sqrt{F_1 F_2 (1 - F_1)(1 - F_2)}} \quad (6.3)$$

is the asymptotic maximum likelihood estimator of  $\rho$ . Confidence limits on  $\rho$  are easily set up as  $\rho$  is a linear function of  $F$ , and  $r$  of  $Q$ .

A special case. If  $F(x, y)$  is a bivariate normal with correlation parameter  $\rho^*$  then, for  $F_1 = F_2 = 1/2$ ,

$$c_1 = c_2 = 1/2, \quad h_1 = h_2 = 0,$$

$$F = \frac{1}{2} - \frac{1}{2\pi} \cos^{-1} \rho^*, \quad 0 \leq \cos^{-1} \rho^* \leq \pi,$$

$$\rho^* = \sin \frac{\pi}{2} \rho = \sin \frac{\pi}{2} (4F - 1),$$

$$\sigma_Q^2 = \frac{F(1-2F)}{2n},$$

and  $100(1-\gamma)$  percent asymptotic confidence limits for  $\rho^*$  are

$$\sin \left[ \frac{\pi}{2} \{4Q - 1 \pm Z_\gamma \sqrt{8Q(1-2Q)/n} \} \right],$$

as  $\sin \theta$  is a monotonically increasing function of  $\theta$  in  $(-\pi/2, \pi/2)$ .

## 7. Generalization

The results of this paper can be generalized to the case of a set of quantiles in samples from a multivariate population possessing a pdf and satisfying certain continuity conditions. From the discussion of Mood (see footnote 2) of the joint distribution of medians in samples from a multivariate population, and from considerations of the generalizations of eq (3.3), it is seen that the asymptotic distribution of quantiles will be multivariate normal. To be more explicit, the joint distribution of a set of quantiles is obtained as a weighted sum of multinomials and multinomials are known to tend to normal under suitable linear transformations on the variates. Thus to specify the asymptotic distribution only asymptotic means and covariances of the variates are needed. However, these can be determined by considering the marginal bivariate distributions.

Then the following conjecture may be stated:

CONJECTURE. Let  $F(x_1, \dots, x_p)$  be the distribution function of  $(X_1, \dots, X_p)$  possessing a pdf  $f(x_1, \dots, x_p)$ . Let the marginal distribution of  $X_i$  be denoted by  $F_i(x)$  and that of  $(X_i, X_j)$  by  $F_{ij}(x, y)$  and corresponding pdf's by  $f_i(x)$  and  $f_{ij}(x, y)$ . Let

$$\alpha_{i1} < \alpha_{i2} \dots < \alpha_{ir_{ij}}, \quad i=1, 2, \dots, p$$

be  $r_i$  quantiles of  $F_i(x)$ . Continuity of the partial derivatives of  $F$  up to order  $p$  at the points  $(\alpha_{1j_1}, \dots, \alpha_{pj_p})$  is assumed for all possible values of  $(j_1, \dots, j_p)$ . Let a sample of size  $n$  be drawn and the values of each variate be ordered so that

$$X'(i1) < X'(i2) < \dots < X'(in), \quad i=1, 2, \dots, p.$$

Let  $k_{ij}$ ,  $j=1, 2, \dots, r_i$ ,  $i=1, 2, \dots, p$  be the integers such that

$$k_{ij}/n \leq F_i(\alpha_{ij}) < (k_{ij}+1)/n.$$

Then the asymptotic joint distribution of the variates  $\{X'(ik_{ij})\}$  is normal with parameters

$$EX'(ik_{ij}) = \alpha_{ij}, \quad j=1, 2, \dots, r_i, \quad i=1, 2, \dots, p;$$

$$\text{Cov} \{X'(ik_{ij}), X'(lk_{lm})\} = \frac{F_{i,l}(\alpha_{ij}, \alpha_{lm}) - F_i(\alpha_{ij})F_l(\alpha_{lm})}{nf_i(\alpha_{ij})f_l(\alpha_{lm})},$$

$$j=1, 2, \dots, r_i, \quad i=1, 2, \dots, p,$$

$$m=1, 2, \dots, r_l, \quad l=1, 2, \dots, p.$$

Here  $F_{i,l}(\alpha_{ij}, \alpha_{lm})$  is to be interpreted in the following way:

$$F_{i,l}(\alpha_{ij}, \alpha_{lm}) = F_{i,l}(\alpha_{lm}, \alpha_{ij}) = F_i(\alpha_{ij}), \quad j \leq m.$$

The author acknowledges the helpful criticism and comments by Dr. Edwin L. Crow on an earlier draft of this paper.

(Paper 64B3-31)